

**Interviewees: CRU\_01 and CRU\_02**

**Name: David Lister (A1), Ian Harris (A2)**

**Role / Title: Research Associate (A1), Researcher (A2)**

**Organisation: Climatic Research Unit, UEA**

**Date: 24<sup>th</sup> April 2014**

**Interviewer: Paula Goodale**

=====

Q: Okay. So to start, if you could just explain to me what your work is here and how it fits into the CRU.

A1: Well, first of all, obviously, neither of us do the same thing all the time. But perhaps the best thing to illustrate what I do in the context of what you're interested in is to tell you a bit about the work towards what we call CRUTEM. This is the monthly temperature, global temperature archive, which you probably have or are vaguely aware of.

Q: Yes.

A1: So again, you see already we're talking about a monthly archive. I mean it's feasible sometimes we get sub-daily data, sometimes we get daily data, but ideally, we want monthly data. So it can mean some work pre-processing to go from sub-daily or daily to monthly.

Q: Where does the data come from?

A1: It comes from a variety of sources, but as Phil was saying earlier on, the main source updated, this is regarding the updating of all series, because, you know, every month there are circulations, there are global circulations which people use with monthly updates on it. Now that's the main mechanism of updating and we no longer really get involved in that. The Met Office, the Hadley Centre, do that. But the main function I have is less systematic. There are countries in the world and groups in the world and different projects, who produce homogenised or specific climate datasets for climate studies. And, you know, again, we were talking about consistency, about we'd like from the same place, from the same micro geographical conditions, micro climatic conditions, blah blah blah. But a lot of these series are actually combined. Compiled, combined, if you like, in a sense that people have put different records together from a particular locality. You know, a station might only move a few metres, and in elevation terms that can be quite important. If you've undulating ground, you know, you might move from one place into a relative frost hollow and this kind of thing. Anyway, that aside, CRUTEM is what the database is called now, or strictly CRUTEM4. This is a paper, I'll give you that.

Q: Okay, thank you.

- A1: That's when we had a big update and added a lot more new series. We were talking about the Russians earlier, for example. The Russians came up with a lot of series that were suitable for our uses a few years ago. More recently the Chinese have. The Canadians are really good at this, the US are really good at this. Norway, Denmark, a lot of countries, and it's increasing, they now produce what they call specific climate series, usually homogenised--, you know what homogenisation means?,
- Q: Yeah.
- A1: So, my job at this time of the year tends to be to go to these different sources, like the Chinese source and the Canadian source, etc, and just grab the recent updates. Sometimes it's just literally a few monthly values, sometimes they've revamped it because they run stuff through algorithms, through homogenisation software, etc. And so they might have suddenly had a big rework and so the series could be significantly changed. This explains the efforts that we did last year. This is in conjunction with the Hadley Centre.
- Q: Right, great stuff.
- A1: So that's not highly scientific, you could understand it.
- Q: Yeah. So how do you access that data? Is it freely available?
- A1: A variety of means. Mostly it's freely available, yeah. The bottom line is now that CRUTEM4 is publicly available. All the series that go into the gridded product, are publicly available on the Hadley Centre website, and that's the rules of engagement now, publicly available. So if anyone offered us a load of series that were for the scientific good but could not be released, we can't use them because they have to be publicly available.
- Q: And what reasons might they not want to make them available?
- A1: Like Phil was saying earlier, like we were discussing earlier, they sometimes just literally want to say there's a monetary to these series and we don't want them to get into commercial hands, or it can be political reasons. A variety of reasons. You know, as someone who works in climate, you often despise these attitudes, but you've got to be able to see why people might not want them to be publicly available.
- A2: The UK Met Office won't give away the hourly data, in the UK, you have to pay for that and it's not cheap!.
- A1: As a country, we're certainly not in the lead in terms of making stuff publicly available.
- Q: And that is because it's got an economic value?
- A2: Yeah, well, because they can, I mean there are people like EDF with wind farms and that kind of thing, for whom weather data at an hourly level is essential for them planning wind farms. They've got the money to pay for it, and that sets the price, essentially.
- Q: So by making it free, they couldn't then command that price from the commercial clients?
- A2: Of course not, because EDF will just go and download it.

Q: Yeah, okay.

A1: So as Phil was saying, for academics in the UK, there's a thing called BADC, British Atmospheric Data Centre, I'm not sure that everything's available but a helluva of lot is, a very large suite of data are available.

Q: Where is that?

A1: I mean you log in and you can download stuff. Once you've registered, it's quite easy, you've been there, Harry, where is it? Oxfordshire?

A2: It's on the M4 corridor, it's not far from Oxford.

Q: Is it owned by a public agency or is it commercial?,

A1: Yeah, it's public,

A2: Yeah, British. I'm trying to think where it is exactly. I'd have been there several times.

A1: British Atmospheric Data Centre.

Q: I'll have a look up anyway.

A1: I mean, it's specifically to serve the academic community. And I guess you can negotiate if you're in the non-academic community, but they might send you to the Met Office. I mean it's Met Office data.

Q: And do you have to use that under certain conditions then?

A1: Yeah, sure. If you were working on a project with someone else, I can't remember exactly, but I suspect that you can only use the data yourself, you can't pass them on to another project or something like this--

A2: Didcot.

A1: Sorry?

A1: Yeah. Even though, you know, you might be working on a very similar thing and you might wish to save someone some effort and time, because it's rather raw data, most of it. Well, most of the land data that I use is rather raw data and it takes a bit of processing, so it can be quite a job.

Q: So what sort of processing do you do with then?

A1: Well, you might want to put it into a daily template, but for example, for example, you might have--, if it was a daily file, you could have three different readings for one day, simply because one reading might be what came off the instrument. And then the next reading might be a QC'd value because they've determined through their QC software that this value is wrong. Or there may have been two or three different values come out of that instrument for the day, or whatever. I'm really not sure, but you can clean the data up,

A2: Or one's been homogenised and one hasn't.

A1: You can get these different values for the same day. And there are flags, obviously, so you have to then take the flag to know which of those values to use. And it's quite complicated.

Q: And that varies by station, or by kind of dataset or?

A1: Yeah, it just varies. It's just a bit random really. It's a bit difficult to put a finger on it, but if you wanted to know more you'd have to ask the Met Office or BADC.

Q: So you work down at the individual station level, do you?

A1: Yes, to some degree. Yes, yeah.

A2: Yes, and produce products from that.

Q: Yeah. So what happens to the data then when it gets in?

A1: Well, it depends. You might download it from a website. Sometimes an agency might send you it in a more convenient format, for you personally. 'Cause, you know, if you're a citizen scientist you might just want one station series for a period of time, but if you're doing what we do, you might prefer to get the whole 100 or something. They might send you it, but it's in a convenient format. But when I say convenient, it doesn't have to be monthly, it might be daily or whatever.

A2: And the format, there are no absolute recognised formats, although netCDF is quite common as a sort of binary format, with the metadata included. But that's not really accessible to citizen scientists that much because you need specialist tools to decode it.

Q: Okay.

A2: So we make all of our data available as ASCII as well and just describe the format unambiguously so that people can access it. But we still, I get people contacting me for the datasets that I manage, because they're having trouble opening them in Excel or something.

Q: So the data that comes from all these different places in the world, there's a lot of variation in it?

A1: Well, in what, in terms of formats, etc?

Q: Yeah, in terms of what's delivered.

A1: Yeah, sure. Yes.

Q: And does the quality vary?

A1: In theory no, but in practice to some extent yes. You know, there are different interpretations-- , I mean, we talk about homogenisation, and it sounds as though it's a specific task. Well, it is but the tools for doing it are not, there are different tools for doing it and different people, different organisations may choose to use different tools. That's not to say that one or other is not right or one's better than the other. We don't get involved in that too much, we just take it that the National Met Agencies or the people involved where it's a project, generally know

more about this particular series than we do. So they're the people who are best suited to do it, in terms of knowledge,

A2: And they will be in their own position, as we are to an extent, though, we used this last year, so we're using it again this year, you know.

Q: I suppose what I was getting at, are they working to a common kind of international standard for the homogenisation, or do they have their own preferred way of doing things?

A2: No.

A1: Not really, no. People go to meetings and people write papers and do all sorts of comparisons, but there is no universal consensus.,

A2: There is currently the thing that Peter's involved with, that's looking at different homogenisation methods and comparing them.

A1: Yeah.

Q: Okay, so they are variable.

A2: That's at the Met Office.

A1: It's a living subject, you know, it's evolving all the time.

A2: I've been doing homogenisation with a particular process that is an algorithm that's quite well known and I coded it into FORTRAN and have been running it. But that made me just recognise how many different ways of doing it there were. Because it is, I mean you're looking at a set of stations and trying to find odd values in individual stations. There's obviously going to be a lot of ways of doing it.

Q: So you've got some variation there and variations in quality and formats. How does that impact on what you do?

A2: Him, more than me.

A1: It makes it more difficult to write a sort of generic processing program. When I say a processing, I mean to get it into a common format. And commonly, I do do a few simple checks, not sophisticated checks, but you can still receive something that isn't quite right.

A2: Duplicate values, values that are ten times what they ought to be, or a tenth of what they ought to be, transcribing errors. I mean, I tend to work from bulletins, so Climat and MCDW, which are gathering together reports from all over the world and putting them in a common format, because I run an automated process that produces the CRU TS datasets and that's completely automated, provided you actually sit there and run it. So that expects particular formats. But you work with individual country formats, don't you, more than I do.

A1: Yeah. And, you know, when we say homogenous, it's quite a complex subject really, it's a very complex subject. But, for example, one of the very simple checks that I do, and these are some I did yesterday, is I just plot the annual temperature series. And it's not uncommon to see a sudden step, and then the line continues. Well that's a fairly good indication. Now I'm

not suggesting there's one there, but there's something a bit weird with that one. Can you see there's something weird at the beginning of that series.

Q: Yeah, it's got kind of a peak.

A1: So that would make me put a query flag on that one, and that kind of thing. So we do still do a few checks.

Q: Yeah. Okay. So do you get them into this gridded dataset?

A1: We put them into, yeah, a kind of format that we use and then the Hadley Centre process. And they make this five by five gridded product which is made available.,

Q: Okay, so you get the data ready and then they process it.

A1: Yeah.

A2: But they're using Phil's algorithms, aren't they?

A1: Yeah. They're using a long established algorithm to do what they do.

A2: But they've coded into C++ I think.

A1: Yeah, they've recoded it, because again it was all partly to do with this business, the one we were being accused of manipulating the data and all this kind of thing, so everything had to become more transparent.

Q: So what's the relationship between CRU and the Hadley Centre?

A1: We're partners in terms of CRUTEM.

Q: Are you equal partners, would you say?

A1: I wouldn't like to say on that, you'd have to ask Phil about that.

A2: That's difficult to say, because there isn't a financial aspect to it.

Q: No, no, no, it's a working relationship.

A2: Well, that's how you'd measure it.

A1: Yeah. And I think to some extent we were, when I say we, it wasn't in my time, but we were the founders of this product.

Q: Yes, that's what I've understood, yeah.

A1: Yeah, and then the Hadley Centre, for various reasons, have come on board as a partner.

A2: Yeah.

Q: So they've clearly seen some value in what you were doing here, would you say?

A1: Well, I guess so.

A2: Probably, but the thing is, we've been in partnership with them for various things forever. I mean when they were moving from Reading or Bracknell, there was a strong move to get them to move to Norwich, but Exeter won in the end.

A1: Another thing is, CRUTEM is a component because if you look on the HADObs website, it's a component into some of their other products. You know, the land and sea product, for example.

Q: Yeah, so that's the HadCRUT, is it, the land and sea combined?

A1: It can be, depending on the name of it, yes.

A2: HadCRUT, yeah. And they always provided the SST. That was going on for some years before the actual action moved to them.

Q: That's the sea, yeah. Okay. So did they ask you to do this or were you doing it already and they said they'd like to work with that?

A1: I think it was we were doing it already but it's a more of a question for Phil is that than for me, I think.

A2: Yeah, I don't know what the politics were of that. I mean, it wasn't something, I don't think it's something that we feel we lost.

Q: Do you see a benefit from working with them?

A2: Oh, absolutely.

Q: Does it add to your work here?

A1: Sure, sure.

A2: Yes, there's a lot of synergy.

A1: I mean it's possible, you see, that given that Hadley were using this product, you know, or at least the input to this product, it's possible that when people were trying to trash our reputation, the Hadley Centre wouldn't want their reputation trashed as well. So they perhaps became slightly more interested in it at that point.

Q: Oh, okay. So yeah, it kind of gave them some sort of auditing--?

A1: Yeah.

A2: Yeah. Quite a few of our ex-PhD students have gone to work there and some of them are still there.

Q: So when you've done all this work here, then it goes to the Hadley and they do more work on it?

A1: Well, let's put it this way--, it's probably in that second paper there. So they will do some analyses and produce some stuff on their website. It's not to say that we wouldn't run the software here, just to check for ourselves, but the stuff that's going out to the public is run at the Hadley Centre.

A2: Yeah, I mean the previous version of CRUTEM was produced in a collaboration between CRU and the Met Office and that was 2006.

A1: And was before any controversies or anything else.

A2: Yeah, and that was Phil's programme running it and me working with him. And we did a lot of work on the station data before it went into the gridding program, to try and clean it up, and produced a lot of statistics about the quality of it. But that's all been completely superseded by the work that went on in CRUTEM4, which had more effort.

Q: So that's quite interesting, though. So how do you validate or evaluate the quality of your product here?

A2: Well again, different answers from different people. I mean mine is not a product where every cell is guaranteed to be accurate, because mine is a complete grid with no missing values, designed for models to use. So I just produce metadata files that show how many stations were contributing to each datum, at each timescale. So people are then free to use those to evaluate the data in whichever way they want. They know whether a cell is just defaulting to the climatology or whether it's got two stations near it reporting, or so on. With CRUTEM it's more a case of knowing that you've done a lot of work on the station data, but they're all compared with other data centres.

A1: Yes, and we tend to rely upon, for example, as I say, the stuff I do that's described in these bits of paper, taking stuff from Canadians, who know their stations, they know their station histories, they're good at doing this kind of thing. And a lot of it is on we're trusting their work. As I say, I do do simple checks, but nothing sophisticated.

Q: So your knowledge of their expertise is important.

A1: Yeah. And at the end of the day, comparison graphics will be produced, and not just for the current time but going back in time. And if there are any changes apparent, in any parts of the world, you might say, "Oh, well, we know why that is because we just received an extra 100 stations from China." Or, it could mean there's an error somewhere, you know.

Q: So you look at it all the time.

A1: Yeah, not just ourselves but people at the Hadley Centre and no doubt the rest of the scientific community.

Q: So the Hadley are obviously one place that this goes to, is there anyone else who uses it in the state that you produce it? Or do they use it when it's been then post-processed?

A1: Not really. I mean the thing is, any scientist can download all these station series from the Hadley Centre for any purpose that they wish.

A2: Which we'll produce--

A1: Yeah. So, you know, they may be looking at something else. It's like Harry was saying earlier, you have people, often biologists, perhaps are trying to relate whether our climate and something else, some biological activity or something, and they're not particularly proficient at computing in the way that we handle data. And they maybe want to use Excel, which there's



nothing wrong with that. But they'll often absolutely bogged down when they get this massive global gridded dataset and they just want one little tiny bit out of it.

A2: And we have to be careful of how much we're dragged into bespoke reformatting for people. I tend to try and avoid it because it's a slippery slope.

Q: So there's a lot of requests but you're--,

A2: Yeah. I mean probably one or two of the requests I get a week would be best met by me spending a couple of hours reformatting everything for them.

Q: And so are they researchers, what kind of people are they?

A2: Yeah, they tend to be working other fields, so economics or as you said, bioinformatics.

A1: Yeah, or social sciences for that matter, you know, adaption to climate change in Africa or something like that.

Q: Or they're academics, generally.

A1: Yeah, or they're academics.

A2: But I don't think we've ever sent them away unhappy, it's just that there's a limit to how much you can divert off what you're doing. I don't think any of us have any time written down in projects to help with random enquiries on the internet.

Q: But do you feel there's some value in responding?

A2: Oh, of course yes, yeah, absolutely. Because well, I mean, I suppose just putting it purely from a selfish point, you know, it's more people are using my data so as I get more citations.

Q: Right, okay, yeah, yeah, because you have to cite data nowadays.

A2: But I wouldn't say I'm particularly bothered about citations myself, it's more a case of just helping people. And I suppose, to be cynical again, you're protecting your reputation and that kind of thing but, you know, what's wrong with helping people?

Q: Yeah. Do you have any interested in what they do with it?

A2: I try not to, too much, because again it's too much data. I don't want to have to handle that as well.

A1: Sometimes they'll actually put you, you know, name you on a paper.

A2: Oh, I've been named in MSc's and things.

A1: There's another issue, and that is we can produce these sophisticated global datasets, but until people start to use them, that's when some errors come to light. So it's something that's slipped through the net here, they might find an error and they say, "What's going on here?" Well, it's in our interest to look at that.

Q: And you get that feedback, do you?

A1: Yes, oh yes.

A2: And that's more common in my area than yours, probably.

A1: Yeah. Yeah.

A2: Because mine's a bit more of a general-purpose dataset.

Q: Yeah, I mean I've been in that situation myself. Yeah.

A2: Mine's a high resolution plug it into models dataset. And we had a situation last year, where I got a set of stations off of David for Mozambique and other countries, and the Mozambique ones turned out to have the latitudes inverted. So it put it in the northern hemisphere instead of the southern hemisphere. Now this is something that I don't check for, because normally I'm getting that data from Climat and MCDW and they've been through the WMO process, so they've all been locked into WMO numbers and lats and longs. And the first I knew of it was when someone contacted us and said, "We've got some very funny values in Algeria, compared to your previous version of your data." And we went and looked, and then David went, "Oh God, I bet I know what that is." And it was a mistake that had been made in another project..

A1: Yeah, mistakes that had been made, I'd seen it, and I ran everything and corrected for it. And then there was something else came about, and then I-- the second rerun I forgot to do the correction, so that's how it--, so it didn't originate from me, but then it was my lack of care on the second occasion that it went through the system.

A2: It was my complete lack of care in checking what you've given me. [laughs]

Q: But you got the feedback and it was therefore--,

A2: Yeah, and in that case I judged that as not justifying a reissue because it only affected parts of Algeria, it didn't really affect anywhere else. It was only Algeria because Algeria didn't have many stations. So in a load of new stations dumped in actually affected the values there. It didn't affect Mozambique 'cause that already had lots of stations. And I've just put an advisory note on the BADC telling people about it, to say that if they want a corrected version, I do have one, I did run one. But that'll be corrected in this year's version, which will be in a month or so.

Q: So we said earlier that these datasets have been available publicly for a while now.

A2: Oh, this has always been available.

A1: Well, that was your gridded product that's been available ever since the year dot.

A2: It's always been available. It's the station data that's newly available because the various countries who'd given us it had various injunctions against us giving it out. And then the Met Office released it anyway, so we [laughs] well, we may as well do it then.

Q: And generally would you say you see it being a positive thing, the data being available?

A2: Yes, absolutely.

A1: Oh yes.

- A2: Yeah, we've nothing to gain from not releasing data. For a start, it means we don't have to constantly field criticism and requests from people. The strange thing is, though, I gather that it hasn't been downloaded very many times since it was made available, almost as if people just wanted to complain about it not being available. But I mean the Met Office will be able to tell you how many times the station data has been downloaded, but there was a lot of fuss about it being available.
- A1: Yeah, but to reiterate, there were real issues, because when the data--, well, first of all, when we started gathering the data back in, I suppose it was in the 1980s, there was no political overtones or anything like that.
- A2: Well, no one was taking climate change seriously.
- A1: No, and it wasn't an issue, you know, it was a global temperature series and it was of interest to some academics, and that was about the end of it. It's when it became involved in climate change and being one of the cornerstones of proving that climate change or whatever was happening, that it then got into the political arena. But there were, definitely, and I was involved in some of the data collection in the--, well, it was in the '90s actually. And we specifically sent, we had a specific sentence that we put in, "We will not release any of these data without your consent." So therefore a lot of the data going in were, on the face of it, confidential. And some donors actually said, "You must not release these data." So there was a real issue there. It was all glossed over during all the pseudo scandal, it was all glossed over, but there was a real issue of whether or not we could let people have these data.
- Q: Because it wasn't your data originally.
- A1: Exactly.
- A2: You do still see a lot of restrictions online on climate data saying it's for academic use only. The Met Office do that a lot.
- A1: Absolutely, yeah. And I've recently done some work using a dataset called ECA&D dataset which is a European project, a long running European project, which is fronted by KNMI, which is a Dutch weather service, climate service. They have a very large amount of Europe climate data on there and there is a percentage that is not available publicly. So they're producing a European gridded product, and they're doing what we used to do, but they aren't falling foul of any criticism. You know, you could say well it's making these gridded datasets more useful to the climate community, but some of the input series are not publicly available. So it's still built into the system (that is, the withholding of some data series).
- A2: We're were having a very interesting conversation at, the project was called EURO4M that's just finished, and the successor is UERRA which is now carrying on. And I was at the sort of final meeting for EURO4M and the beginning of UERRA which was at the Met Office a few weeks ago. And a very interesting conversation with one of the guys from Brussels about the fact that really, a European project, it ought to be mandated that all the countries that are

involved in it ought to release their data just to the project. Because what we have is a set of the analysis models in this project and when they show what stations they're working from, as a map of Europe, the coloured in area is always their country. Because modellers from one particular European country have access to the entire synoptic network of their country's reporting stations at an hourly level, but no one else does. So their model is heavily weighted towards their country, plus whichever countries are free, and other countries have a block, you know. Norway have just made theirs available in the last couple of months, so that's now filtering through. But there is a lot of parochialism and again, as we said, it's down to economics as much as anything.

Q: So how would you say your dataset that you produce here compares with others that are done in--?

A1: [Pause] It's hard to say I don't know, Phil's a better person to answer that question.

A2: Well, CRUTEM and CRUTS are both compared in their relative papers with other datasets. I mean if you've got the CRUTEM paper there, that does compare it with, what, GPCC, and others.

A1: Yeah, there are other products.

Q: Is it sort of the coverage or that--,

A1: Yeah, I think the best thing to say is they are relatively similar. But there are differences and some of them not tiny differences. But at the end of the day if you see the output, you know, compared to each other, they're not sensitive to the finer detail because they tend to give a similar result.

Q: And is that similarity interesting in any way, does it prove anything?

A1: Well, I'm not qualified to say that, but I would say yes, it does. It proves that the methods of doing these things are not that sensitive in some respects.

Q: Because you're using different methods?

A1: Well, you see, how they tend to work is, or at least ours works, is that we have what we call a standard period, and so in our case it's still 1961 to 1990, for various reasons. And so we produce what we call climate normals for the period 1961 to 1990. And then all of the monthly values before and after that are related to that standard period, and we express the values in anomalies, the anomalies with respect to the 1961 to 1990 climatology. And that has to happen. I mean you could on to another period, but that basically has to happen.

A2: I mean that was originally chosen because that's where the bulk of the station reporting was. But I mean obviously over the 20th century the numbers of stations that are coming online were increasing and that sort of peaked in about 1980, around that, in the '80s, and started to decline again as budgets got cut. I mean this is anomalies of temperature from CRUTS and these are compared with the University of Delaware temperature series which [shows data on computer screen]

Q: Okay, so you're working across the baseline sort of.

A2: And so you can see these are all anomalies.

Q: Yes, above and below that kind of--,

A2: And the '61 to '90 bit will always average to zero. And this is an example of what we do, this is a comparison country by country, or this is region by region but we do country by country as well, of CRU against University of Delaware, for CRUTS, not for CRUTEM. And so University of Delaware is in a pale grey, which pops up here and there, and we're in black. And we do that kind of thing, but there are also statistical variants, differences between variance maps, that kind of thing.

A1: You might like to have a look at the website, it's an initiative that's being going on for a few years now, called the ISTI, International Surface Temperature Initiative.

A2: Which is the one I referred to earlier, with Peter's, you know--,

Q: ISTI?

A1: Yeah.

A2: Yeah, it's at the Met Office. No, it's not at the Met Office, but it's--,

A1: Well, it's really centred on the US, but the idea was hatched at an international meeting at the Met Office a few years ago. I am actually involved in it. But it's trying to go back to first principles as far as possible. In other words, it would like to receive data, original data, in any form, hard copy, scanned or whatever, and go back to basics and they've made good progress on it. So it's worth having a look at that website.

A2: Surfacetemperatures.org.

Q: Surfacetemperatures.org, okay. And they're taking from global stations, are they?

A1: Yeah.

A2: And they've got a very good level of access too.

A1: Yeah. And there's a fair degree of data rescue involved in it as well. Because, as we were saying earlier, there's still an awful lot of data, older data mainly has not been digitised.

Q: Okay. So is it right, so you're working with new data here, what comes in.

A1: No, not necessarily, because--,

Q: Do you do a lot with the older?

A1: For example, if the Canadians send me, erm, when they tell me they are finished for this year, I will download the whole series back to when they start, so it could be in 1880 or something. And I will then, before I put them into the archive, I will then just check the differences between that and the previous version.

Q: So sometimes they've gone back further?

A1: Sometimes they've gone back further, they may have--,

A2: Or they've done some QC on it.

A1: Yeah, they may have used a slightly different algorithm to do the combination of the different series, all kinds of things can happen.

Q: Do you have any kind of opinions on the data rescue or the--,

A1: The more the merrier, I think most of us would say that. You know, it's something that has to some extent been neglected and time is running out in some situations. And we put great value in older data, you know, it's part of the picture. You need the older data as well.

Q: What makes good older data, for you, for your work?

A1: We need the antecedents of it. We like to know what we call the metadata, and you would soon get a feel of the quality of the data. And ideally, if it was within the terms reference of some organisation like the Met Office or something, some national organisation which had rules and regulations, because again when you're going back to data collection I mean you soon find out, for example, if temperature is not being recorded within an approved screen. The temperatures don't mean anything really, you know, they just get absolutely crazy values.

Q: So you need to know that it's been collected correctly.

A1: Correctly.

A2: But, yeah, I mean what we're looking for is--, I mean going back temporally is important, but plugging in spatial gaps is more important, especially as we deal with gridded products, they rely on there being stations everywhere.

Q: So what are the challenges in that respect?

A2: Well, the Sahara, obviously, there's not too many stations there anyway. But there are parts of, well there were parts of South America but actually the Amazonian Basin got filled in quite nicely a couple of years ago.

A1: Yeah, I mean it still doesn't go back as far as we would like, but yes, there are big, big, big inroads have been made in recent times.

Q: And is that coming from stations or from data rescue?

A2: Data rescue, I think, mainly.

Q: Archive or records.

A1: Some of it, yeah. Because I mean, you know, we've got this old Colonial history.

A2: A lot of it's scanned in, but they need to be OCRd. They might be in the supplementary information, but I don't think it's in this one. [Pause] Oh, well that's coverage by geographic area, in terms of numbers of stations. So the black is the stations we use and this is temperature, the [inaudible 0:36:32] temperature range. So the temperature's at the top, you see North America has this massive rump of stations. But again, you can see this curve, this

is time, along this axis. So you can see the numbers of stations going up and then trailing off and that's pretty much the same in all the other regions. So like Europe here, it's up and down. USSR, same. There's more precip stations.[shows data on computer screen]

Q: So why would the number of stations reduce?

A2: Because they're closed down.

Q: Literally because they're not running anymore?

A2: Yeah.

A1: It's often purely economic. Sometimes, like we were saying earlier with the breakup of the Soviet Union, some of the independent States were no longer able or willing to release their data, it can be things like that.

Q: So changes in boundaries?

A1: Yeah, politically, changes.

Q: So another thing that we've noticed is that the Met Office is now collecting lots of amateur data. Is that of any use to you guys?

A2: Yeah, as long as the Met Office are happy with it. [Laughs]

A1: Yeah.

Q: Until we talk to them in detail, we're not quite sure why they have this.

A2: Well, the thing is that you go back more than about 150 years and it was a gentleman's pursuit. You know, if you had your nice house in the country, you had a weather station.

A1: Well, also, it was the educated people in society, so it was the doctor, the vicar, and people, and then perhaps Lord somebody or other who owned the estate, they were the sorts of people who ran the weather stations.

Q: And nowadays it's more your computer scientist--

A1: Not necessarily.

A2: Well, the local Met Office.

A1: Not necessarily.

A2: Although other people do it, but the Met Office is, because most countries have got a network that they're happy with-- most of them.

A1: I think it can be useful because a lot of these amateurs stick by the rules. You know, by the nature of it, they are keen to be accurate. There are organisations, there's an organisation which just escapes me at the moment, but it's run by a guy at Reading, and that's purely for amateurs. But most of them, they know the rules, they know they have the correct instruments, they know the rules. For example, there's nothing wrong with increasing the density of observation, particularly in terms of rainfall. And sometimes, if there was an extreme event, particularly a rainfall event, it's not very likely that there'll be a major station,

Met Office approved station, right at the epicentre of this event. And they really do like to focus in on these major events, like these, you know, the serious floods in the southwest, at Boscastle was it, and places like that, if they can get more information, if they've got reliable amateurs in that kind of situation, their information can be invaluable, absolutely invaluable.

Q: And you wouldn't rule it out for the temperature sets as well?

A1: Not at all.

Q: No?

A1: No, not at all.

Q: Is that a phenomenon that's worldwide, or is it just UK specific?

A1: [Pause] I don't know.

A2: It's probably more UK and Europe than anywhere else.

A1: I think in this country it's fair to say that for temperature we are reasonably well covered by the official stations. It's a more conservative element than say a precipitation. You know, within a space of a couple of miles you could get a massive difference in rainfall out of an extreme event.

Q: Okay. So-- [pause] So what do you see happening with this work in the future, will it change much or has it got to a place where it's--,

A2: The types of projects that we get involved in have changed, I suppose. I mean, the breadth of science relating to climate change has changed and-- Apart from the fact that a lot of mitigation and adaptation has crept in, and there's a lot of projects looking at that. But I mean the project I'm involved in with UERRA, and before that EURO4M, is looking at better prediction of extremes in Europe, droughts and floods and wind and so on. And they're running all these reanalysis models to try and look at common traits of how they represent extreme-- So they need a lot of reliable observations for that and I'm looking at--, do you know, ECMWF at Reading?

Q: It rings a bell.

A1: It's European Centre for Medium Range Forecasting, and whatever the W is, Weather.

A1: I think it's Medium Range Weather Forecasting or something like that.

Q: Right, okay.

A2: They run a well-respected reanalysis model called ERA. You know what a reanalysis model is, it's like a regular model but its got sets of observations, so it doesn't run off just on its own internal data. And I've been working on there, the observations that they use, which is a huge library of hourly observations, and I'm trying to clean those up. And so that isn't a project that would've happened ten years ago, I don't think.

Q: So you think reanalysis is something that's going to--,



A2: Yeah. Reanalysis seems to be coming up a lot more now. I mean modelling all goes on. I've been involved in projects where we were providing observations, and modellers, to test their models against. I've been involved in providing model data to people. I've been downloading it, reformatting it for them and creating stats and giving them that. So, essentially it's all data work, that's the thing.

It's becoming more detached from the station. I suppose we are moving further away from the raw and more into the processed and analysed. But you're just being stretched thinly between the two.

A1: It's true to say that funding for this type of work has probably diminished in recent years, because people often wanted to move on, and we had enough existing data and they wanted to move on and push the money into development of models and sophisticated--,

A2: And [inaudible - both talking at once]

A1: Yeah, yeah, sure.

A2: And from funding bodies perspectives, it's all the same pot.

A1: So it's a bit more difficult now. I mean it's still ongoing, because some of the stuff Harry has just told you about, it's a bit more difficult now to keep the old sort of rescue, data rescue going. And the other thing is, the international climate is still a bit negative in some respects, from a political perspective or whatever. And the parent body, the World Meteorological Organisation, which no doubt you've heard of, in my opinion they're not quite pushy enough. But then, it's all a diplomatic exercise and it doesn't push you. Being pushy doesn't work in diplomatic circles really. But what I'm trying to say is it can be frustrating sometimes that there are there attitudes still out there that, you know, "We do not want to share our data," for whatever reason.

Q: So you see it as being something that they need to--,

A1: I think it's that, in conjunction with a push towards strengthening data rescue activities, etc, they're actually quite important.

Q: So, more cooperation, more data rescue?

A2: Yes, absolutely.

A1: But, you know, all you need is a civil war somewhere in Africa for five years and that's it. It can be just wipe the sheets clean for five years or something like that.

A2: My absolute dream for a long time has been 10,000 or 20,000 completely automated weather stations in half length containers, solar and wind, whatever, completely impervious, just dropped at grid points all over the surface of the planet. [All laugh] Marvellous.

Q: But is that ever going to happen?

A2: And then you just send it back a 100 years. Brilliant.

Q: Yeah, time machine.

A2: No, it won't ever happen, the politics wouldn't allow it.

Q: So what keeps you motivated and interested in this?

A1: Well, I've always had an interest in weather and climate. Originally it was an amateur interest in weather, meteorology. I've always been interested in climate issues.

A2: I've always been interested in environmental issues, so I'm motivated to be part of this and to help. And I'm reasonably good with numbers and data and programming and that kind of thing, so it's a niche.

Q: So do you see your work producing value to different kinds of communities, to people?

A2: Yeah, absolutely.

Q: What kind of people are they? Who do you think benefits from what you do?

A2: Well, principally it's anyone who uses the data. So, people doing catchment modelling of river systems, town planners, that kind of thing, up to countries planning power policy, that kind of thing. And then, a lot of users of the project or other projects either in the climate sphere, or in adaptation or mitigation spheres, or other spheres, and it supports all of those as well, so it's an absolute myriad of uses.

Q: Is that a motivation for you, or are you just motivated by the fact you like working with the numbers?

A2: Well, like I say, helping people.

A1: Yeah, and we're not just, we describe to some routine stuff that we do, but we're not doing that all the time, you know.

A2: We're not doing anything all the time.

A1: No. We get involved in lots of other things.

Q: So what's the thing that you enjoy doing the most, David?

A1: Well, I guess it's just being part of the climate scene, that's what I'm interested in. And it is something I have a personal interest in anyway.

Q: And is that internationally, or is that here at CRU?

A1: Yeah, sure. Yeah. Oh, we're very international here, very international.

A2: Yeah, it is nice interacting with people from other organisations, from other countries. It doesn't happen very often live. I mean I get to go to the occasional project meeting but they're infrequent. But yeah, it's nice that there's a sort of band of people all over the planet, working on using the same nomenclature and trying to understand the same things. And they all know the limitations of what we're doing and what position we're at in terms of real knowledge, and it's really knowing that you're not acting on your own in this.

A1: And I think there's another thing, that a lot of people outside, especially the critics, think we're all sort of evangelistic environmentalists who are frantically trying to change society and everything else to try and save the planet.

A2: Well, I am. [Laughs]

A1: But for me, I just want to get to the truth, and it's reliable datasets, the best we can do at this point in time, that's my real interest.

Q: So it's the quality and, yeah.

A1: You know, we shouldn't be showing too many leanings in that respect, we have to be unbiased as scientists to do what we do.

Q: So what about the science, is there an element of the science that makes it exciting work?

A2: Well, I suppose. It's difficult to say, because it's so politically charged and it's in the papers every day and this sort of thing. It's hard to evaluate how I would feel about it in a frictionless vacuum kind of situation. Whether I could actually say, "Am I interested in this science for the sake of the science?" I don't know. I got this job by accident.

Q: Right? [All laugh] How did you get the job?

A2: I was coming to do an MSc at the University Environmental Impact Assessment. I'd just left BT, having worked there for a number of years, and I wrote to all the schools looking for summer work. And Trevor Davies, who was the director of CRU then, rang me up and said, "We can offer you some summer work." And I came and did some work here on--, they'd done a downslope wind study, katabatic winds, these are the cold winds that flow down slopes and kill the grapevines in France. And they did a study in Wales, famous for its grapevines, [all laugh], with a couple of masts on a hill measuring all the winds. And I did some work in I think MatLab to try and visualise that, which I'd never used before. And then they advised me not to do the MSc and to work for them instead, so I stayed here, and that was 18 years ago. [Laughs] I couldn't say I aimed for this [inaudible - all talking at once].

Q: No, but you found your place, yeah.

A2: Absolutely.

Q: That was interesting. Is it common here that you've got people who've been here for a long time?

A1: Particularly in the recent past, I mean there's quite a few people who have retired recently. But this was a very solid core of people, with people coming and going at the edges, a very solid core of people for a very long time.

A2: Yeah, I mean when I joined, as I say the entire building was full and there were four or five people over in a room in ENV. The room across there had four staff in it; that was the general office. So they weren't doing science they were doing admin and finance, and that sort of thing. We had our own finance person, and a secretary and director, all that kind of thing. And it's all gradually been whittled down. I mean although the climate-gate thing did

accelerate that very slightly, most of that had already happened by the time--, we'd already lost financially independence before that.

A1: Yeah, for various reasons. I mean I guess when this unit stood largely alone, it was very vibrant and it was like plenty of income appearing. The fashion's changed slightly and the income starts going into a slightly different direction, like mitigation, adaption, then if the incomes here fell a bit then we'd have probably struggled to have our own admin team, etc.

Q: So you've got this core of people, have you--?

A1: It's not so strong now. I mean it's still relatively strong.

A2: Yeah, the worrying thing is we're probably part of the core now.

A1: Several of the long established members have retired or moved on to other centres.

[section deleted from published transcript at request of interviewee]

Q: So do you think that kind of history has given your work a particular direction?

A2: It gives it a strength and a foundation.

A1: And also, you've got people here who have a great deal of experience and they can nearly think of a way around any problem, whatever you're trying to do, because they've got so much experience of doing the--, maybe the methods have changed, but the actual concepts are still the same.

A2: Yes. The methodology is sound, absolutely.

Q: Does it change over time, the methodology, does it incrementally change?

A2: Well, for new datasets, yes, you would because you take advantage of latest theories, or the purpose of the dataset. But for things like CRUTEM, the algorithm has stayed the same for decades.

A1: Yeah. Well, computing power has increased significantly. I mean at one time we might have had a certain format for your dataset, which itself was space saving but which made it more difficult to work with. Now we've been able to relax some of those constraints.

A2: Someone wrote to me yesterday, or last week, and I got it yesterday, and they were having trouble with one of our legacy datasets, although it is still current, because it's in the old format, which is a space saving format. Although it's a gridded dataset, it lists each cell, so you get the lat and long of the cell, and then the values for that cell, and then the next cell. So although it's a gridded dataset it's not, and you can't load it into Excel and get a grid. It's all the coordinates are on each line, and I say to people, "You know, I'm sorry, everything has been formatted, it's a legacy dataset, that's how it is."

Q: So it's more computing power than science that's changing your methodology at the moment?

A2: Yeah, the change of data storage has made a huge difference to things.

A1: Yeah, storage and computing power.

Q: Okay. So you can work with more data?

A2: Well now I've got about three terabytes in there and another three backing it up every day, and that would have been unthinkable ten years ago never mind 30. It's more storage than the entire building had ten years ago.

A1: Yeah, you know, you might have to work several seconds to do a simple search on a relatively, what we would we say now was a relatively modest size database. Now, you think nothing of it, you know. And you used to worry about having lines of data in that were all missing values, but now it's far more convenient to have the missing values there and then everything's complete.

A2: Yes, because when you zip it up, they all vanish anyway.

A1: Yeah. And back in the old days, they had to try and economise on that kind of thing. It was a harder job.

A2: Yeah. Even FORTRAN programs, the old ones, have really, really short variable names, to save space. Again, every byte counted.

Q: Yeah, I've seen that, yeah.

A2: I remember, you know. I do remember that. And that's accelerated ahead of our need for it.

Q: Right. So now you've got the power that you need, for a while.

A1: But we're not running sophisticated climate models like they'll tell you about at, you know, at the Met Office.

A2: There's been debate over the years as whether we would do modelling, but the most we've ever got to that is MAGICC, which was a hemispheric model, so it just treated two hemispheres as two items and [inaudible 0:54:34] any interactions between them. That was Sarah Raper, wasn't it?

A1: Oh, probably, I don't know.

A2: And what's his name, the Australian--?

A1: Oh, you mean--,

A2: A director.

A1: You mean [pause] Tom Wigley.

A2: Tom Wigley, yeah. That was their pigeon--. So we've never really done much modelling here. But we could now, with the processing power we've got. You could do some simple models, it's just that enough people are doing them, we don't need to.

Q: So your with data niche--, that's your expertise?

A2: Yeah, it's observations plus data processing of other people's outputs.

Q: That's great.

A2: Everyone's a link in a chain.

Q: Yeah, absolutely, and we're trying to follow that chain which is--,

A2: Yes, quite.

Q: Okay. Anything else you think I might need to know?

A1: I don't know. I think there's a link, there's a nice link here that might not be relevant [shows web page], but it tells you of a data rescue, a happy story of a data rescue initiative. And if you're interested in--, I mean, it doesn't directly relate to us, but it's about the rescue of some data that were in a museum in Punta Arenas.

A2: Is that Clive related?

A1: No. I think Phil mentioned it earlier. IEDRO, I-E-D-R-O, so it's--, I could email it to you if you want the link. How do you want it?

Q: I'm sure Google will find it.

A1: You'll find it.

Q: Yeah. Yeah, that's great.

A1: It's Chile, IEDRO, museum, that's the key word, museum, data rescue.

A2: But lots of archives are rotting away in basements and things, you know, and need to be--,

A1: It's just an aside is this, but a key agency in scientific observation was the Catholic Church at one time, the Jesuits, the Jesuit branch of the Catholic Church. They were keen to make good observations and everything, and they were often in the less developed parts of the world, but they were often not so keen to make them share the data with other people, if you understand me. But there are still people chasing Jesuit datasets.

A2: It sounds like a terrible Dan Brown novel, doesn't it?

Q: And they're buried away in church archives and things.

A1: Well, some are. I'm not an expert on it, but particularly I think South America. I mean there was a French guy who spent a few months here, and he was specifically sort of data rescue and he reckoned he had good contacts with the Jesuit groups. But they're not very open in terms of embracing people to come in and, you know, "Here's what we've got, and looking at our precious library," and all this kind of thing. Having said that, a lot of inroads will have been made. And this could have been in a similar principle, I don't know whether it was a Jesuit thing, because a lot of this Jesuit stuff goes back further than this, this just went back into the 19th century, a lot of it's earlier than that. The early sort of missionaries, if that's the right term, I don't know.

Q: Okay. Brilliant.

[END OF INTERVIEW]